NASA-CR-198010

INTERIM
IN-52-CR
8 CIT.
45513
P. 27

Annual Progress Report

Period Covered: 5/1/94-4/30/95

NASA Cooperative Research Agreement
NCC2-542

Psychophysical Evaluation of Three-Dimensional Auditory Displays

Frederic L. Wightman, Ph. D., Principal Investigator
Waisman Center, University of Wisconsin
1500 Highland Avenue
Madison, WI 53705

(NASA-CR-198010) PSYCHOPHYSICAL EVALUATION OF THREE-DIMENSIONAL AUDITORY DISPLAYS Annual Progress Report, 1 May 1994 - 30 Apr. 1995 (Wisconsin Univ.) 27 p

N95-24564

Unclas

G3/52 0045513

Introduction

This report describes the progress made during the first year of a three-year Cooperative Research Agreement (CRA NCC2-542). The CRA proposed a program of applied psychophysical research designed to determine the requirements and limitations of three-dimensional (3-D) auditory display systems. These displays present synthesized stimuli to a pilot or virtual workstation operator that evoke auditory images at predetermined positions in space. The images can be either stationary or moving. In previous years, we completed a number of studies that provided data on listeners' abilities to localize stationary sound sources with 3-D displays. The current focus is on the use of 3-D displays in "natural" listening conditions, which include listeners' head movements, moving sources, multiple sources and "echoic" sources. The results of our research on two of these topics, the role of head movements and the role of echoes and reflections, were reported in the most recent Semi-Annual Progress Report (Appendix A). In the period since the last Progress Report we have been studying a third topic, the localizability of moving sources. The results of this research are described below.

The fidelity of a virtual auditory display is critically dependent on precise measurement of the listener's Head-Related Transfer Functions (HRTFs), which are used to produce the virtual auditory images. We continue to explore methods for improving our HRTF measurement technique. During this reporting period we compared HRTFs measured using our standard open-canal probe tube technique and HRTFs measured with the closed-canal insert microphones from the Crystal River Engineering Snapshot system.

Detailed Progress Report

1. Localization with Moving Sources

An important requirement of a usable 3-D auditory display is synthesis of veridical auditory image movement. Sound image movement, defined as a change in the direction of a sound relative to the listener's head and ears, occurs even when the sound source itself is stationary. In a natural situation, listeners move their heads, and these movements cause a change in the position of a stationary source relative to the listener's head. The changes in relative orientation result in predictable changes in the spatial cues produced by the sound source at the listener's ears. Such changes could be important since in theory they can provide essential information to the listener about source position. We have found that listeners judge the position of both real and virtual sound sources more accurately if head movements are encouraged. Using a virtual auditory display system, we presented sound sources which appeared to be stationary to the listener by coupling the image synthesis to the listener's head position in real time. The listeners were encouraged to move their heads during the stimulus presentation. Front-back reversals often reported by some listeners when localizing virtual sources disappeared and judgments of source elevation were more accurate. The details of this experiment were presented in the Semi-Annual Progress Report (Appendix A).

The results of the first experiment on head/image movement do not address the question of whether the improvement observed in localization performance requires proprioceptive feedback from actual head movement or auditory image movement alone. Since 3-D auditory displays are likely to find application in situations in which the operator's head may not be free perception) can be obtained with source movement alone. It is possible to provide the listener with changes in the acoustical cues similar to those that accompany head movement simply by moving the source, while the listener's head remains stationary. There is very little published data on listeners judgments of apparent position of a moving source. Previous research on source movement has focussed either on listeners' ability to judge "time to contact" of a moving source or on the minimum angular movement that is detectable. We are currently conducting experiments in which listeners are asked to localize moving sources and in which listeners are allowed to move the source to aid localization.

Using the "absolute judgment" paradigm described in our publications and previous progress reports, we tested listeners in several conditions in which the stimulus was a moving source. The first condition did not provide a "naturally" moving source but simulated movement with static sources. It consisted of presenting 3 250 msec noise bursts that changed either in azimuth or elevation by 10 degrees. An example of an azimuth change would be a sequence of 3 sources at 50, 40, 30 degrees azimuth and 20 degrees elevation. An elevation change might consist of 3 sources at 160 degrees azimuth and -30, -20, -10 degrees elevation. This condition served to provide contextual information, without actually simulating a naturally moving source. Since we were primarily interested in how this condition would affect the resolution of front-back confusions, we only tested four listeners who made front-back confusions when judging the position of static virtual sources. The listener's task was to report the azimuth, elevation and distance of the last (third) source in the sequence. None of the listeners appeared to benefit from the additional cues provided by this condition. Listeners' performance in this task was remarkably similar to their performance in the static source condition. Figure 1 shows the results from a single listener in the static source (left panel), azimuth "movement" (center panel) and elevation "movement" (right panel) conditions.

In a second experiment, we presented listeners with a virtual source that moved 40 degrees in azimuth. The stimulus was a noise burst 1 sec in duration and the rate of movement was 1 degrees/25 msec. In one condition the listener reported the apparent starting position and in a second condition, the apparent ending position. We tested 7 listeners, the 4 listeners that participated in the first experiment and 3 listeners who do not make confusions. When listeners were presented moving sources, their judgments of starting (or ending) source position were no more accurate than their judgments of static sources. Front-back reversal rates in the moving source task were similar to the rates observed in the static source experiments. Data from the static and moving source conditions are presented for two subjects in Figures 2 and 3.

In the third experiment, listeners were presented a virtual source and encouraged to move the source by pressing keys on a computer keyboard. Both azimuth and elevation movement was possible. The stimulus was a dei noise that played continuously until terminated by the listener. Preliminary data suggest that when the listener is allowed to control the source movement, the apparent difficulties that some listeners experience in resolving front-back differences disappear, just as they did when head movement was encouraged. The results from a single listener in this condition are presented in Figure 4. An analysis of the source movement histories indicated that the angular movement was about 5 degrees for both azimuth and elevation for listeners who do not typically make front-back reversals and about 40 degrees for azimuth and 20 degrees for elevation for listeners who do make front-back reversals.

2. A Comparison of Open-Canal and Closed-Canal HRTF Measurements

The fidelity of a 3-D auditory display is critically dependent on accuracy with which we can measure the listener's Head-Related Transfer Functions (HRTFs) that are used to produce virtual auditory images. If the HRTF measurements are not made carefully, or if a generic set of HRTF measurements are used, the fidelity is compromised, often resulting in large increases in front-back confusions and degradations in the perception of source elevation. Currently, we measure HRTFs using an open-canal probe microphone system (Etymotic ER7-C). If the tip of the probe tube is place at the eardrum and the probe remains stable during the measurement session, this technique produces very accurate representations of both the directional and non-directional components of the HRTF. This techniques does have several disadvantages, however. First, it is sometimes difficult to place the probe tube near the eardrum because of the shape of the earcanal. Second, the probe tube microphone is relatively insensitive and noisy. Third since the canal is open, the signal level cannot exceed 75 dB to avoid contamination by the acoustic reflex. Because of the last two problems, averaging is required to obtain an acceptable signal-to-noise ratio. If HRTF measurements are made using a closed-canal insert microphone system, the microphone (a more sensitive one) is positioned at the canal entrance and the signal level can be higher, obviating the need for extensive averaging, since the earcanal is blocked. A potential disadvantage is that canal entrance measurements may not capture all of the directional characteristics of the HRTF.

Six listeners participated in an experiment designed to compare HRTF measurements made with open-canal probe microphones (Etymotic ER-7C) and closed-canal insert microphones (from the Crystal River Engineering Snapshot HRTF Measuring System). During a single session, measurements were made at 126 spatial positions using both microphone systems. The measurements were repeated several times on a different days.

In order to compare the measurements made with the two systems, we find it useful to decompose each individual HRTF into the product (in the frequency domain) or convolution (in the time domain) of two transfer functions. One represents the "average" response of the ear (at the eardrum) to sounds from all directions, and the other represents the departures from that average that are specific to each individual direction. The first we call the "diffuse-field" estimate (DFE), which formally is the response of the ear to a diffuse sound field. The second we call the "directional transfer function" or DTF. The DTFs are estimated by dividing each HRTF by the DFE. Figures 5 and 6 show the HRTF, DFE and DTF at a single source position from two listeners. the solid curves show the measurements taken at the eardrum with the probe-tube system and the dashed curves show the measurements taken at the entrance to the closed ear canal. While the two systems produce very different HRTFs and DFEs. the DTFs are very similar.

Multidimensional Scaling Analysis was used to summarize DTF differences between the two measuring systems and repeatability of each system. The levels (dB) in non-overlapping critical bands were determined for each DTF. The difference between any two sets of DTFs was represented by the Euclidean distance metric, the square root of the sum of squared dB differences. A 29 x 29 matrix was constructed, representing the differences among all 29 sets of DTFs (there were 2 or more sets of DTFs for each measurement system from each of the 6 listeners). This matrix was subjected to the scaling analysis which produced a 3-dimensional solution, accounting for 90% of the variance in the data. A 2-D projection of the 3-D scaling solution is shown in Figure 7. The letters refer to different listeners, with uppercase representing the canal entrance measurements and lowercase representing the probe measurements. The differences between the two systems appear to be no greater than differences among repeated measurements on a given listener for each system alone. For 3 of the listeners, variability among the sets of canal-entrance measurements was somewhat greater than for the probe measurements.

We also evaluated the potential utility of the closed-canal system for measuring HRTFs that can be to produce virtual auditory targets in a localization task. Two sets of virtual sound sources were synthesized, one from HRTF data obtained using the standard Etymotic probe tube system and one from data obtained with the CRE closed-canal system. In both cases the source was a single 250 ms burst of white noise presented over high-quality headphones at about 70 dB SPL. Each of the 126 virtual positions were randomly presented 5 times. Listeners judged the apparent positions of both sets of virtual sources, those made from closed-canal measurements and those made from eardrum measurements. Results from two listeners are shown in Figures 8 and 9. Data from the canal-entrance condition are shown in the left panels and data from the probe-tube system are shown in the right panels. The fact that the patterns of judgments are nearly identical for both sets of virtual sources suggests that the CRE closed-canal HRTF measuring system can be used effectively in the process of producing virtual auditory targets. Its main advantages over the conventional probe-tube system are a much higher signal/noise ratio (thus, shorter measuring time) and less discomfort for the listener.

Publications

Papers

Wightman, F. L. & Jenison, R. L. (1995). Auditory Spatial Layout. In W. Epstein & S. J. Rogers (Eds.), Handbook of Perception and Cognition. Volume 5: Perception of Space and Motion. Orlando, FL: Academic (In Press).

Wightman, F. L. & Kistler, D. J. (1995). Factors affecting the relative salience of sound localization cues. In R. Gilkey and T. Anderson (Eds.), Binaural and Spatial Hearing. Hillsdale, NJ: Erlbaum (In Press).

Macpherson, E. A. (1994). On the role of head-related transfer function spectral notches in the judgment of sound source elevation. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe. NM. (In Press).

Zahorik, P. A., Kistler, D. J., Wightman, F. L. (1994). Sound localization in varying virtual acoustic environments. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In press).

Abstracts

Jenison, R. L. (1994). Radial basis function neural network for modeling auditory space. Journal of the Acoustical Society of America. 95 (Pt. 2), 2898.

Jenison, R. L., Wightman, F. L. and Kistler, D. J. (1994). Self-organizing model of auditory maps. Abstracts of the 17th Mid-Winter Meeting, Association for Research in Otolaryngology, 70.

Wightman, F. L., Kistler, D. J., & Andersen, K. (1994). Reassessment of the role of head movements in human sound localization. Journal of the Acoustical Society of America, 95 (Pt. 2), 3003.

Wightman, F. L., & Kistler, D. J. (1994). The importance of head movements for localizing virtual auditory display objects. Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In Press)

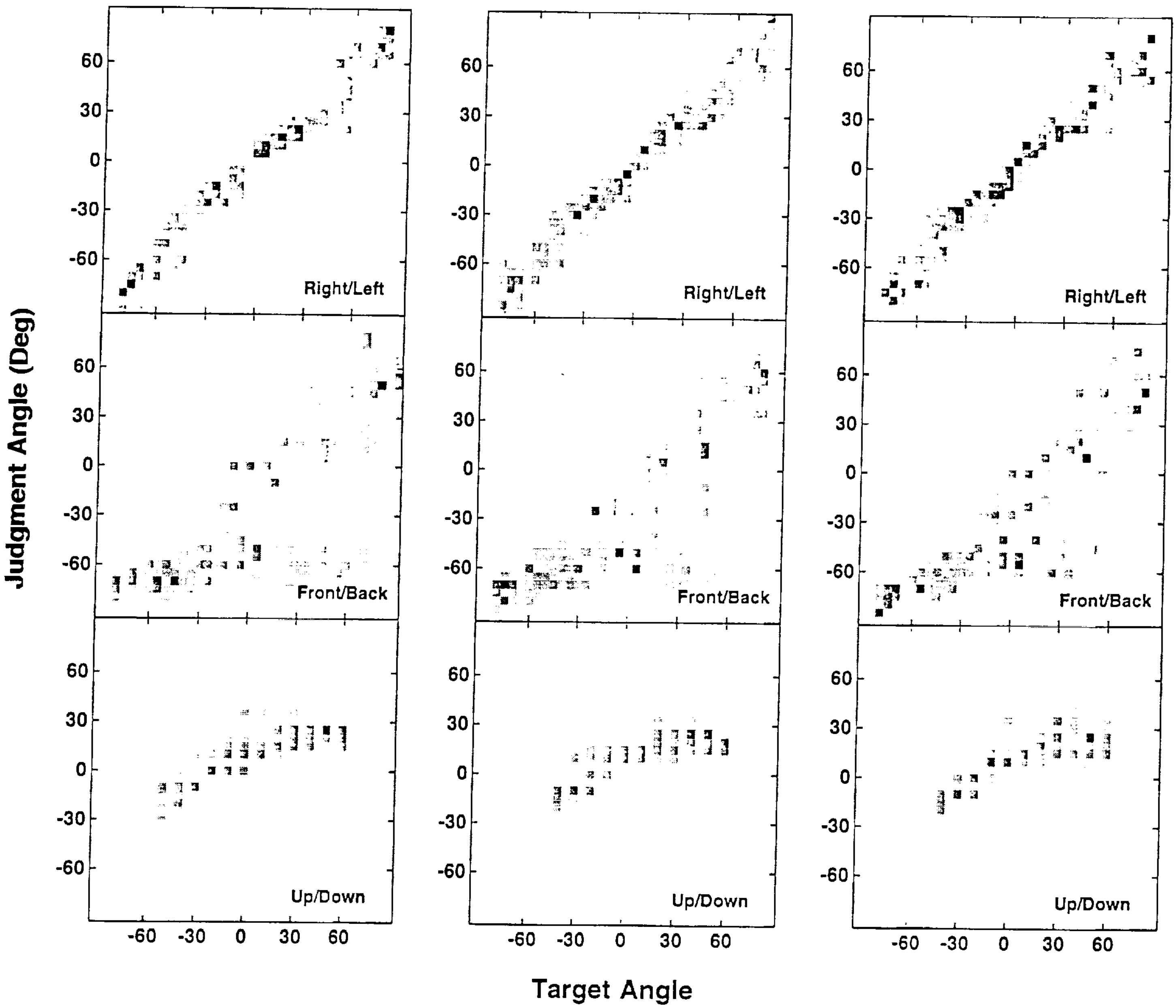


Figure 1. Judgments of apparent position of virtual sources from Listener SMQ in the static source (left panel), azimuth "movement" (center panel) and "elevation" condition (right panel).

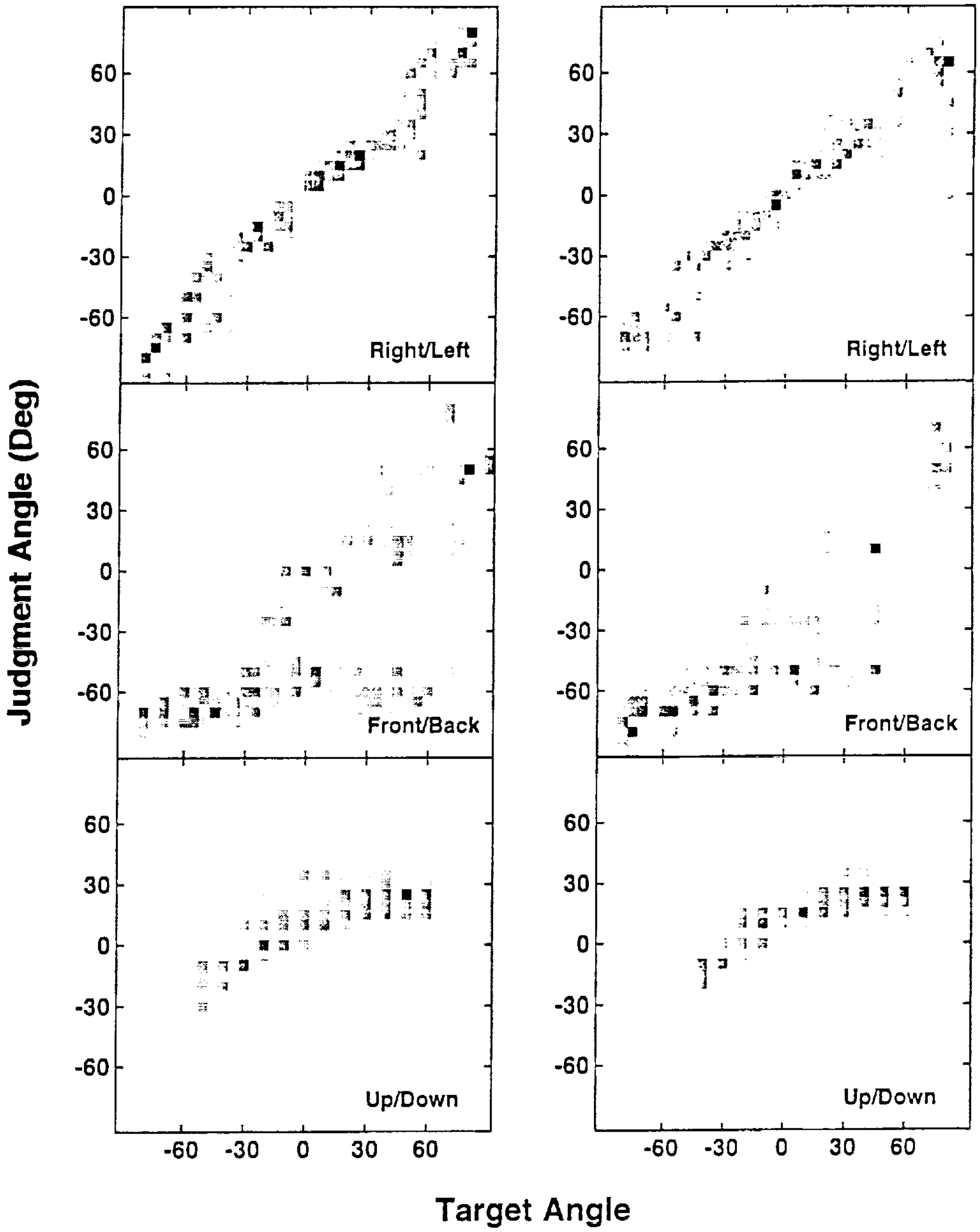


Figure 2. Judgments of apparent position of virtual sources from Listener SMQ in the static source condition (left panel) and the moving source condition (right panel).

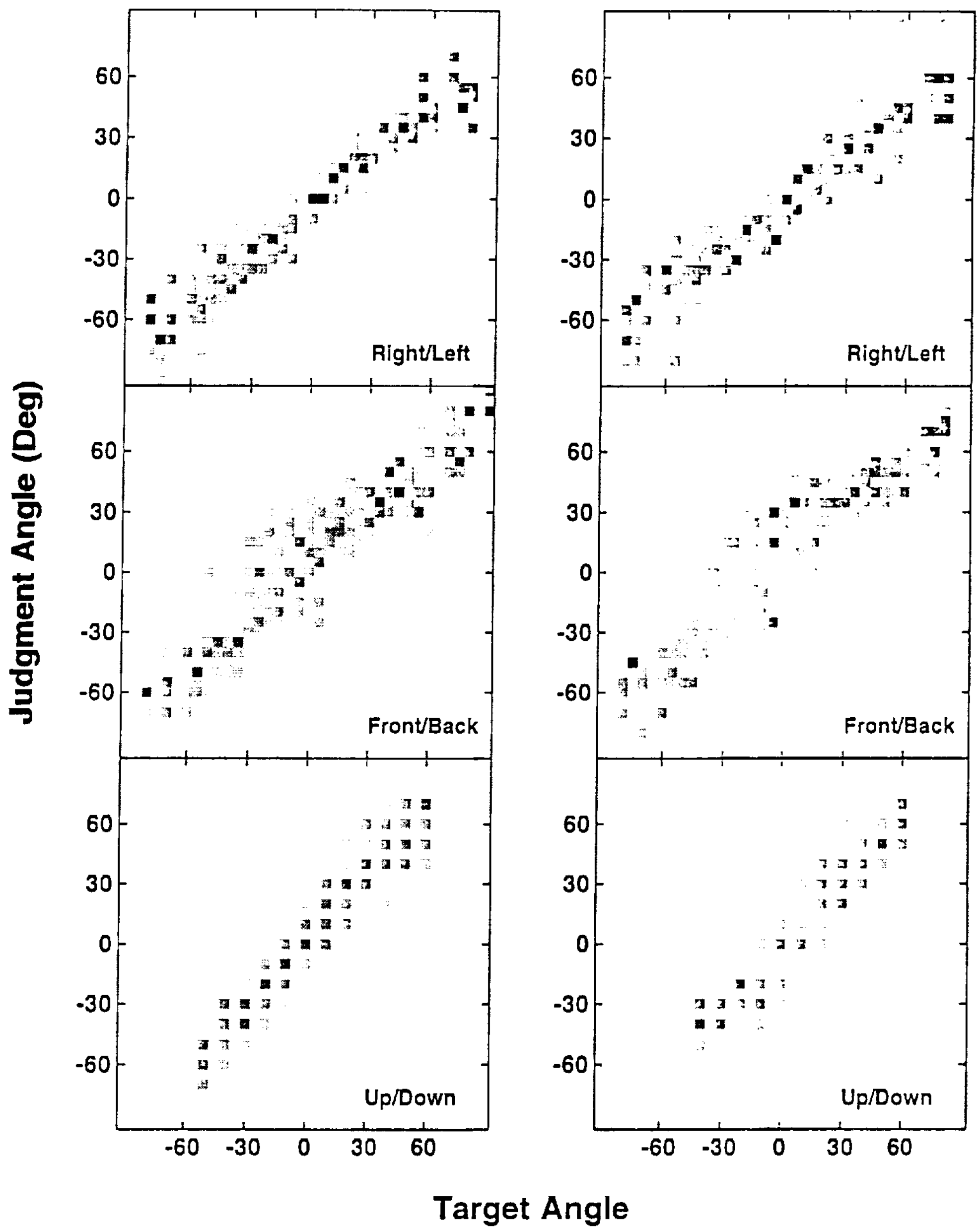


Figure 3. Judgments of apparent position for virtual sources from Listener SNJ in the static source condition (left panel) and the moving source condition (right panel).

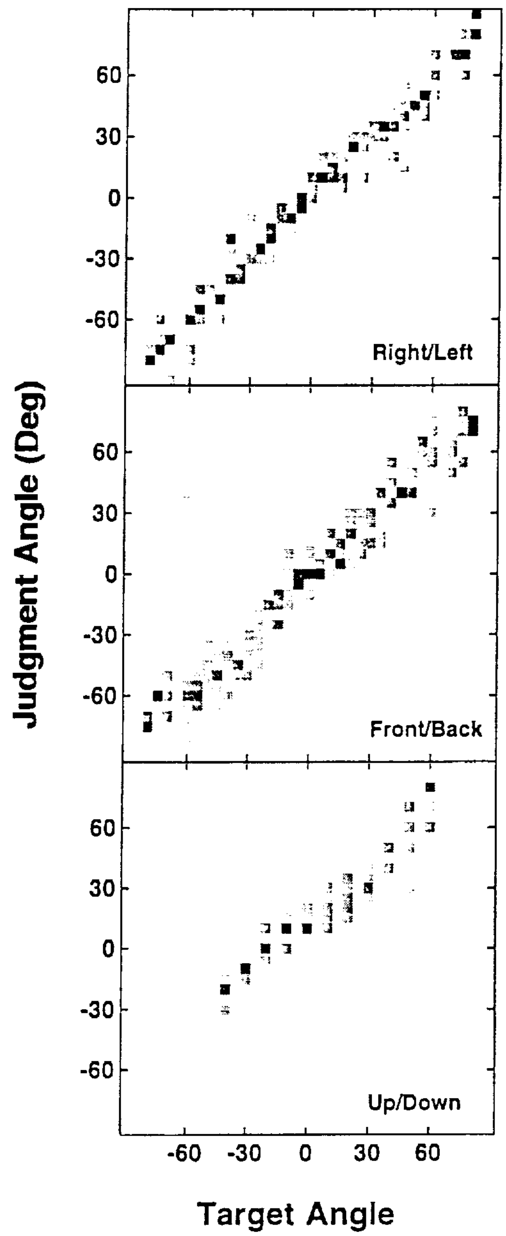


Figure 4. Judgments of apparent position of virtual sources from Listener SMQ in the condition in which the listener controls the movement of the source.

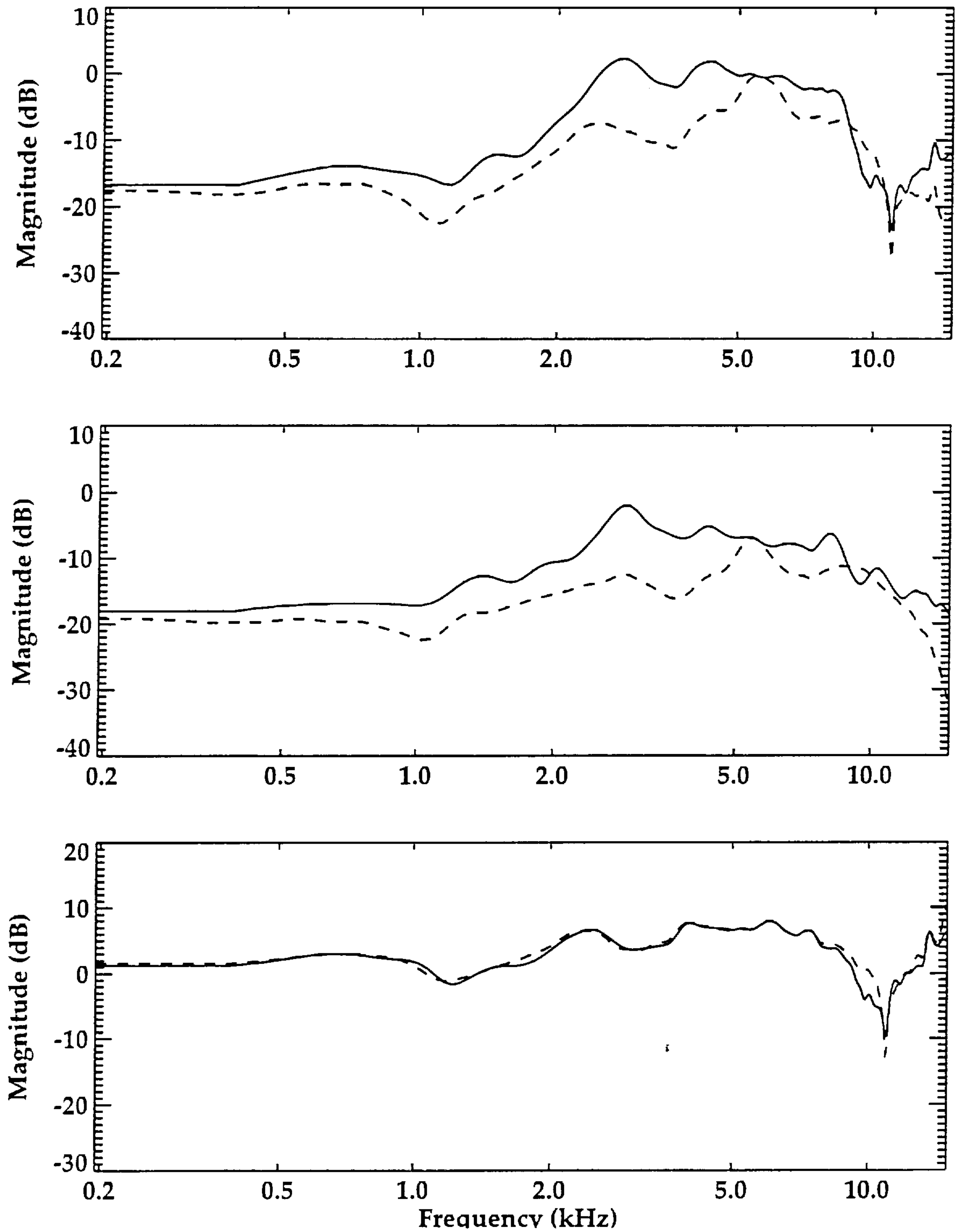


Figure 5. The top panel shows the raw HRTF magnitudes for a single source position from Listener AFW. The measurement obtained with the probe microphone is plotted with a solid line and the measurement obtained with the canal entrance microphone is plotted with a dashed line. Diffuse field estimates are plotted in the center panel and directional transfer functions are plotted in the bottom panel.

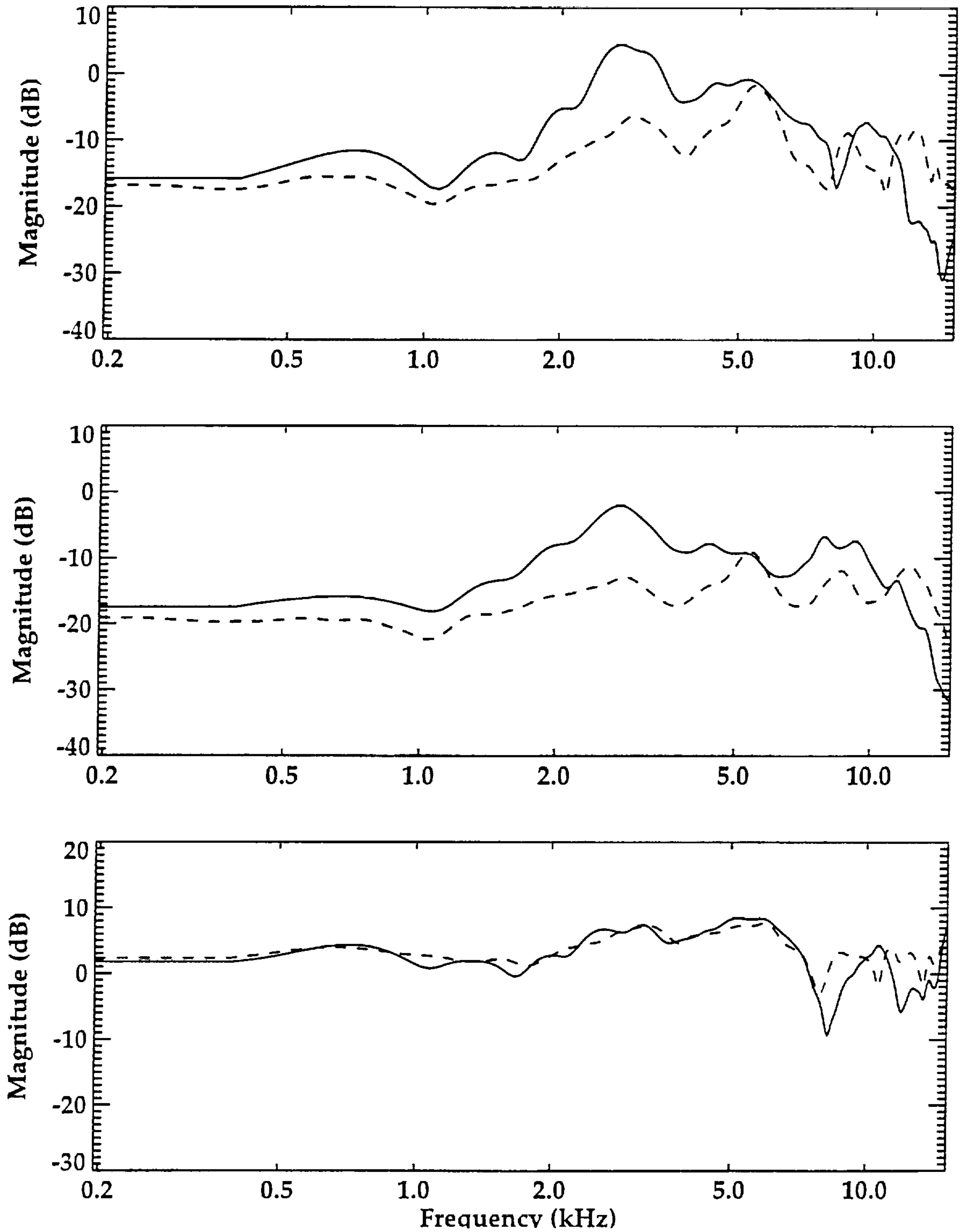


Figure 6. The top panel shows the raw HRTF magnitudes for a single source position from Listener SNF. The measurement obtained with the probe microphone is plotted with a solid line and the measurement obtained with the canal entrance microphone is plotted with a dashed line. Diffuse field estimates are plotted in the center panel and directional transfer functions are plotted in the bottom panel.

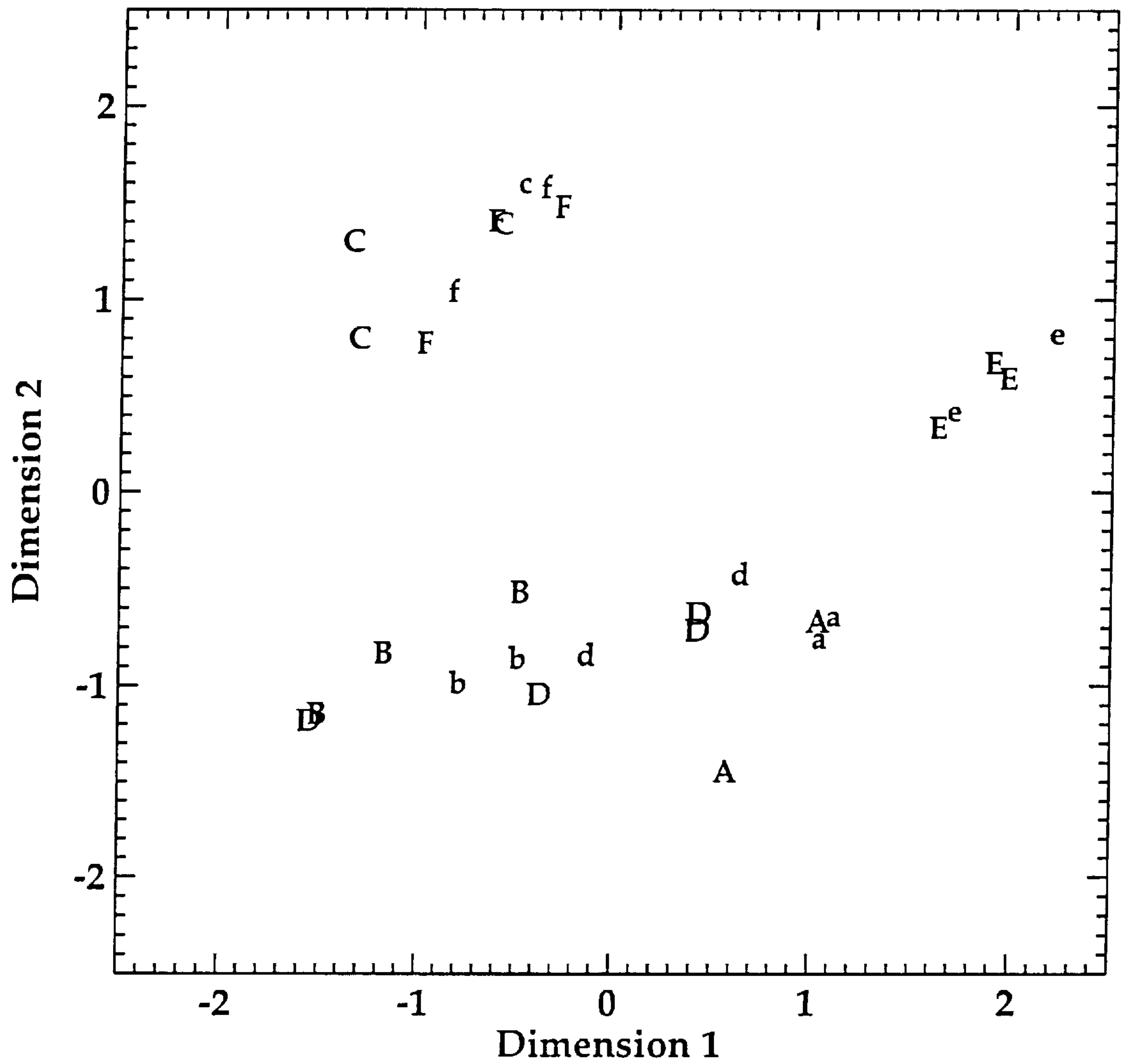


Figure 7. Two-dimensional projection of the 3-dimensional Multidimensional Scaling solution of DTFs estimated from measurements made with open-canal probe microphones (lowercase) and closed-canal insert microphones (uppercase). Each listener is represented by a different letter.

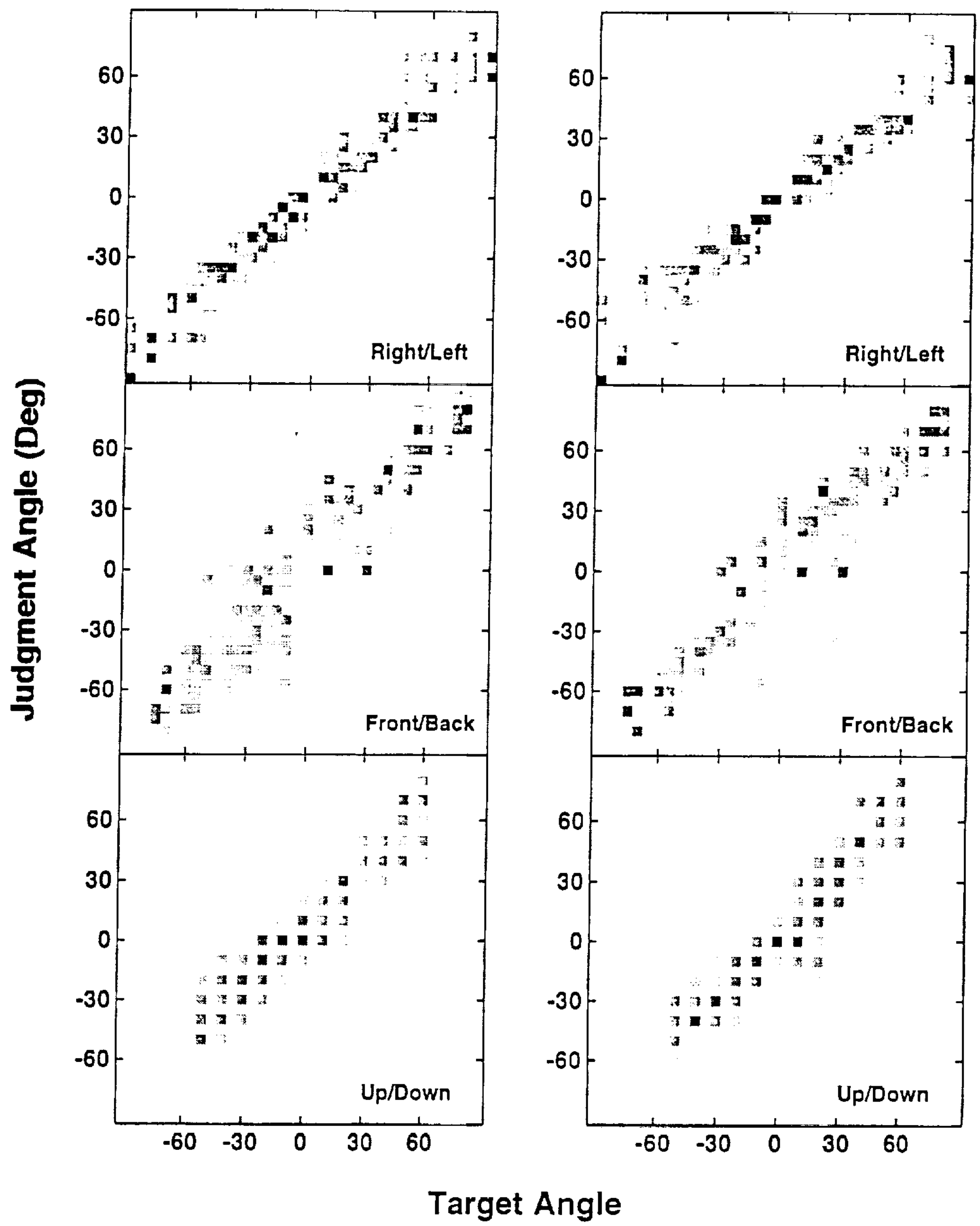


Figure 8. Judgments of apparent position of virtual sources produced from HRTF measurements with the open-canal system (left panel) and with the closed-canal system (right panel) from Listener SNJ.

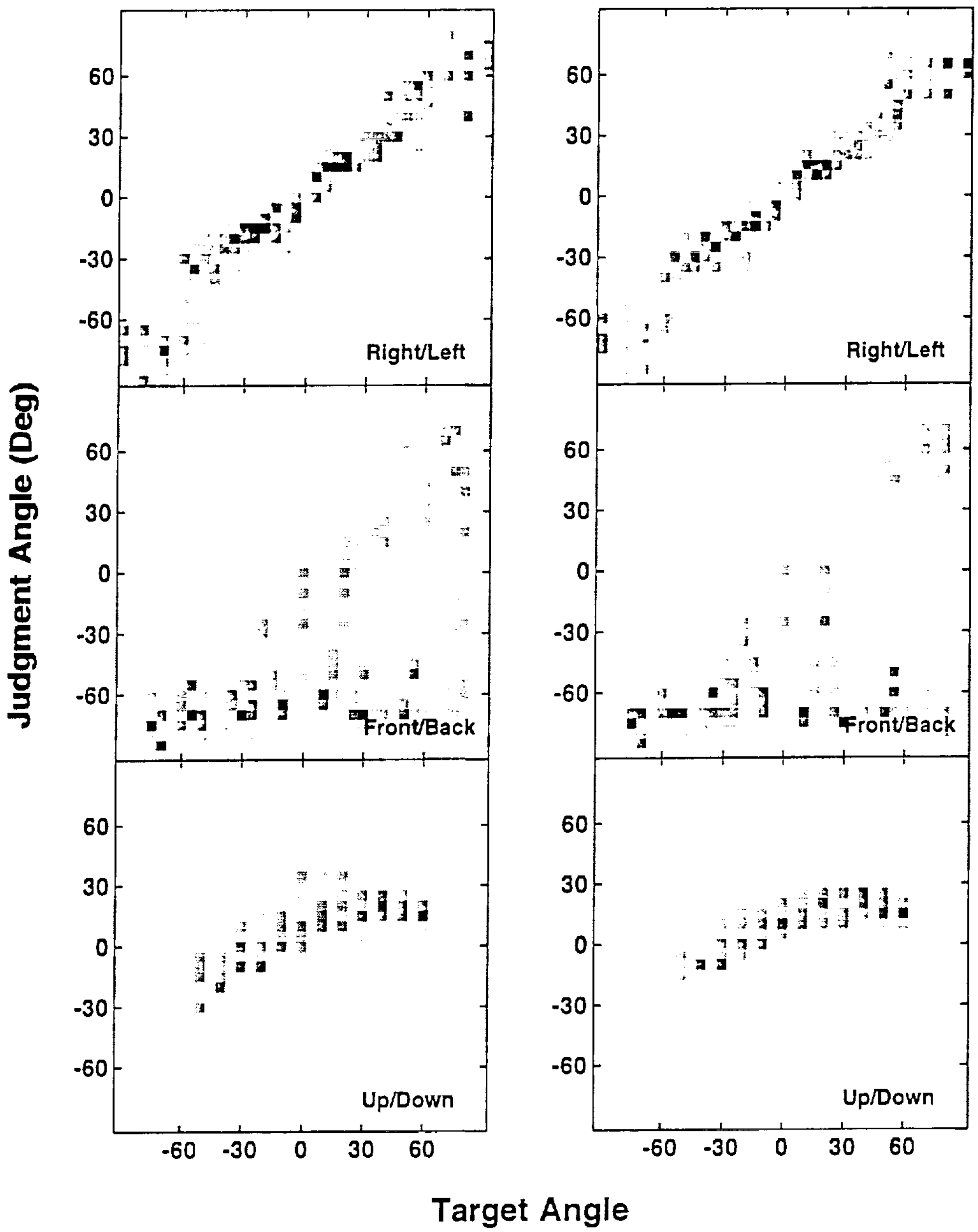


Figure 9. Judgments of apparent position of virtual sources produced from HRTF measurements with the open-canal system (left panel) and with the closed-canal system (right panel) from Listener SMQ.

APPENDIX A

Semiannual Progress Report

Period Covered: 5/1/94-11/1/94

NASA Cooperative Research Agreement
NCC2-542

Psychophysical Evaluation of Three-Dimensional
Auditory Displays

Frederic L. Wightman, Ph. D., Principal Investigator
Waisman Center, University of Wisconsin
1500 Highland Avenue
Madison, WI 53705

Progress Report

The fidelity of current virtual auditory display systems is limited primarily by the occurrence of front-back confusions and poor representation of target source elevation. Work during this reporting period attempted to achieve a better understanding of the importance of several acoustical cues that we believe are important for achieving high quality front-back and elevation perception and good externalization with virtual auditory displays. Experiments were completed on the role of dynamic cues provided by head movements and on the role of cues provided by echoes. Additionally, we continued our efforts to relate spectral features of HRTFs to perceived sound source location by formulating a model which attempts to predict elevation judgments from the frequency of the primary spectral notch in the HRTF.

1. Role of Dynamic Cues

When a listener's head moves while listening to a stationary sound source, the interaural time, interaural intensity and pinna cues change in accordance with the head movements. In an experiment described in a previous progress report, we presented 5 listeners with stationary virtual sources synthesized with the Convolvotron, which was coupled to a magnetic head tracker. The listeners were encouraged to move their heads to facilitate localization. Only one of these listeners made large numbers of front-back confusions in the baseline condition in which no dynamic cues were available. The results suggested that the cues provided by this listener's head movements could eliminate these confusions.

During the present funding period we sought to replicate this result in a second experiment with 8 new subjects, 6 of whom made front-back reversals in the baseline virtual source and in the freefield conditions. In addition to the baseline condition in which stimuli delivered to the headphones were not influenced by the movement of the listener's head ("restricted" condition), there were two movement conditions: 1) listeners were encouraged to move their heads to aid localization ("freestyle" condition); 2) listeners were told to point their noses at the sound source ("compulsory" condition). The stimuli were 2.5 s virtual sources synthesized by the Convolvotron using HRTFs measured from each listener's own ears. The position of the listener's head was tracked and the synthesis of the virtual source was modified in real time, in accordance with the head movements to simulate a stationary external source. For those listeners who made frequent front-back reversals in the baseline condition, reversal rates were near zero in the two head movement conditions. We also observed some improvement in perceived elevation, especially in the "compulsory" condition. Data from the three conditions are shown for 2 listeners in Figures 1 and 2.

Analyses of the trajectories of the listener's head movements revealed that while the tracks were idiosyncratic, they were remarkably consistent from presentation to presentation for a single listener. In general most listeners appeared to orient toward the source in the "freestyle" condition. An examination of some of the trials on which the listeners made reversals revealed that the listeners did not attempt to move their heads on the majority of these trials. The 2 listeners who did not make reversals in the baseline condition showed very little head movement in the "freestyle" condition.

Figure 3 illustrates trajectories of head movements in the "freestyle" and "compulsory" conditions for a listener who makes frequent front-back reversals in the "restricted" condition. The four panels show head movement trajectories (indicated by the dotted lines) from four trials on which the same virtual source was presented. Note the consistency in the trajectory on the four trials. Also plotted on the figures are the nominal position of the virtual source, the mean judgment made in the "restricted" condition and the judgment made on each trial in the "freestyle" condition. Figure 4 shows trajectories on two identical trials from a listener who makes few front-back confusions. Note that in the "freestyle" condition, this listener's head movements were very small.

The results strongly suggest that head movements are a natural and important component of localizing sounds and that auditory displays that incorporate head-coupled synthesis will provide a more realistic listening environment.

2. Role of Echoes

An important feature of natural listening environments is the presence of echoes and reverberation. There is anecdotal evidence that suggests that echoes may enhance the externalization of virtual sounds and that they may provide additional cues for resolving front-back ambiguities. In our first experiment, described in a previous progress report, we presented virtual sources that were synthesized to include not only the direct sound but also the first-order reflections off the four walls of an 8 x 8 x 3 m room. Reflections were attenuated by 6 dB to mimic "soft" walls. Listeners' azimuth and elevation judgments were indistinguishable from their responses to virtual sources with no reflections.

In our recent work on this topic, we tested 5 new listeners in three types of virtual stimuli: 1) "dry" virtual sources containing no echoes, 2) echoic virtual sources synthesized using the image model to predict spatial position, time delay and amount of attenuation for the first 20 reflections occurring in time after the direct source path, and 3) "perturbed" echoic sources synthesized with 20 reflections for which the time delays and attenuation factors were computed according to the predictions of the image model, but the spatial positions were chosen randomly. Listeners performed similarly in all three conditions. The details of this experiment are in a manuscript included with this report.

3. Role of Spectral Notches

There is considerable evidence to suggest that low-frequency interaural time difference is the primary determinant of perceived laterality or the "left-right" component of a sound source. It is widely believed that monaural spectral cues are important determinants of the other two dimensions of apparent source position, "front-back" and "up-down" or elevation. However, the nature of the relationship between spectral features of an HRTF measured for a particular sound source and apparent source position is not known. The most prominent features of HRTF magnitude spectra are the high-frequency notches. An examination of our HRTF data indicates that the frequency of these notches changes in a fairly systematic fashion with changes in source elevation. The pattern of change differs across azimuths and across individuals. Consequently, we sought to determine if these differences in notch frequency pattern could be used to predict elevation judgments.

A simple model was formulated which predicts that perceived elevation is determined by the frequency of the primary high-frequency notch in the HRTF of the ear closest to the source. The primary notch frequency was determined "by eye" for 132 positions spaced 30 degrees apart in azimuth and spaced 10 degrees apart in elevation (elevations ranged from -50 to +50). The model further predicts that the variability in elevation judgments is related to the notch frequency gradient such that the steeper the gradient, the lower the variability. Results from an analysis of the variability of freefield elevation judgments of 6 subjects do not support the single-notch model. We conclude that perceived elevation must depend on additional spectral features. The details of this work are provided in an attached manuscript.

Publications

Papers

Wightman, F. L. & Jenison, R. L. (1995). Auditory Spatial Layout. In W. Epstein & S. J. Rogers (Eds.), Handbook of Perception and Cognition. Volume 5: Perception of Space and Motion. Orlando, FL: Academic (In Press).

Wightman, F. L. & Kistler, D. J. (1995). Factors affecting the relative salience of sound localization cues. In R. Gilkey and T. Anderson (Eds.), Binaural and Spatial Hearing. Hillsdale, NJ: Erlbaum (In Press).

Macpherson, E. A. (1994). On the role of head-related transfer function spectral notches in the judgment of sound source elevation. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays. Santa Fe. NM. (In Press).

Zahorik, P. A., Kistler, D. J., Wightman, F. L. (1994). Sound localization in varying virtual acoustic environments. In G. Kramer (Ed.), Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In press).

Abstracts

Jenison, R. L. (1994). Radial basis function neural network for modeling auditory space. Journal of the Acoustical Society of America, 95 (Pt. 2), 2898.

Jenison, R. L., Wightman, F. L. and Kistler, D. J. (1994). Self-organizing model of auditory maps. Abstracts of the 17th Mid-Winter Meeting, Association for Research in Otolaryngology, 70.

Wightman, F. L., Kistler, D. J., & Andersen, K. (1994). Reassessment of the role of head movements in human sound localization. Journal of the Acoustical Society of America, 95 (Pt. 2), 3003.

Wightman, F. L., & Kistler, D. J. (1994). The importance of head movements for localizing virtual auditory display objects. Proceedings of the 1994 International Conference on Auditory Displays, Santa Fe, NM. (In Press)

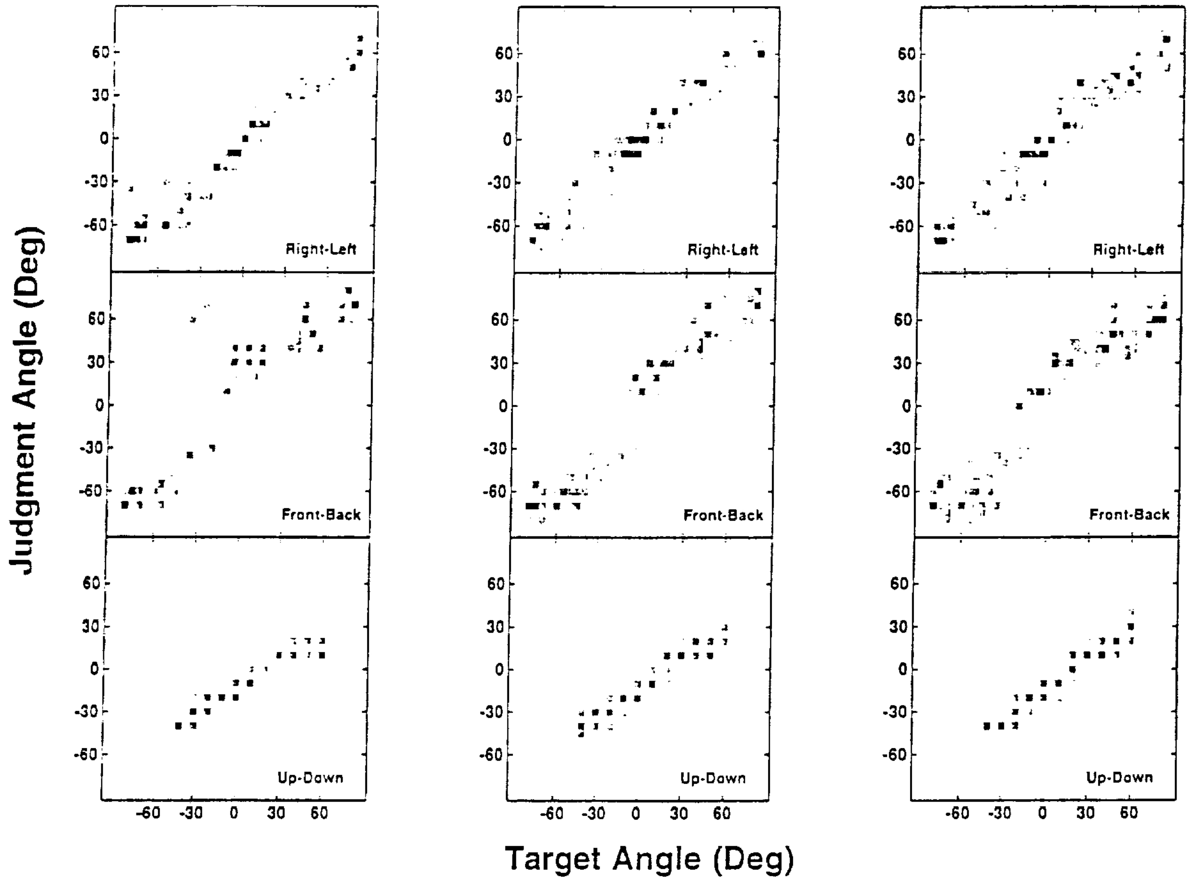


FIGURE 1. Data from Subject SNF in the three head movement conditions: "Restricted" (left panel), "Freestyle" (center panel), and "Compulsory" (right panel).

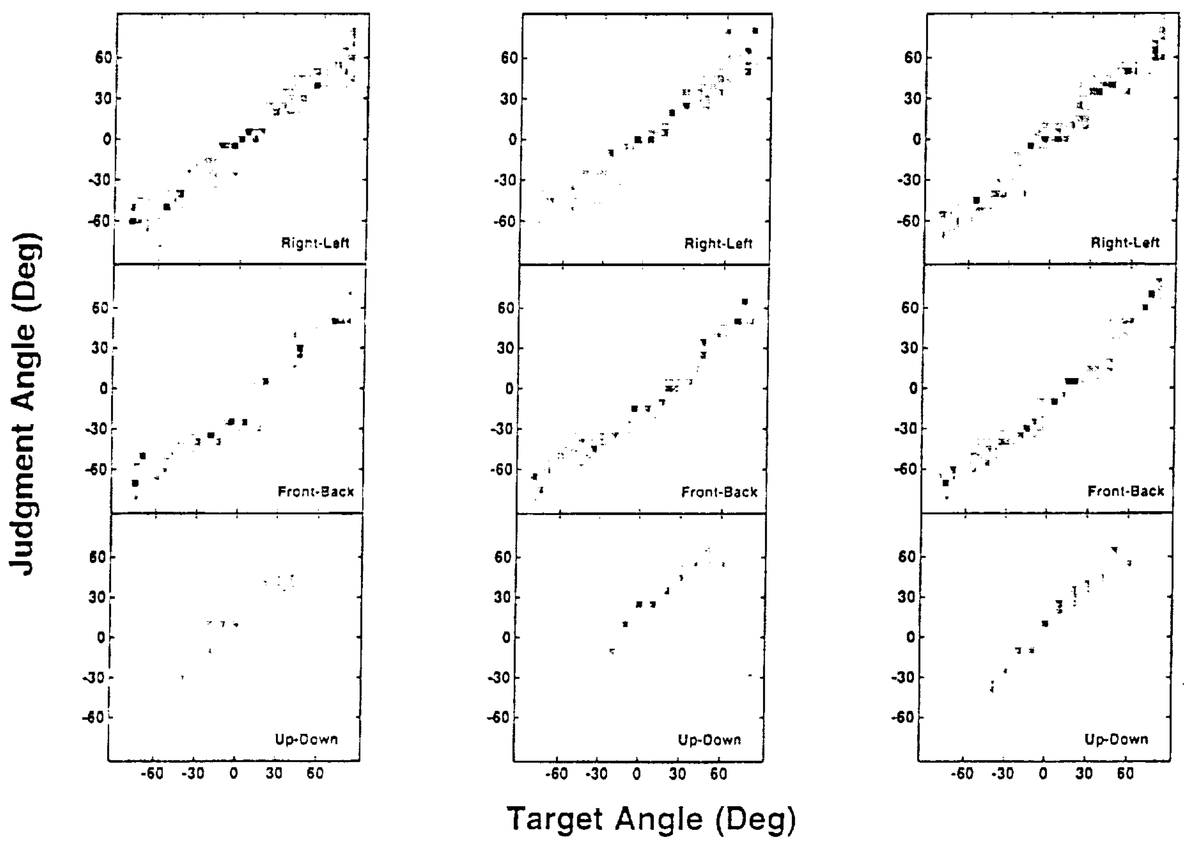


FIGURE 2. Data from Subject SNR in the three head movement conditions: "Restricted" (left panel), "Freestyle" (center panel), and "Compulsory" (right panel).

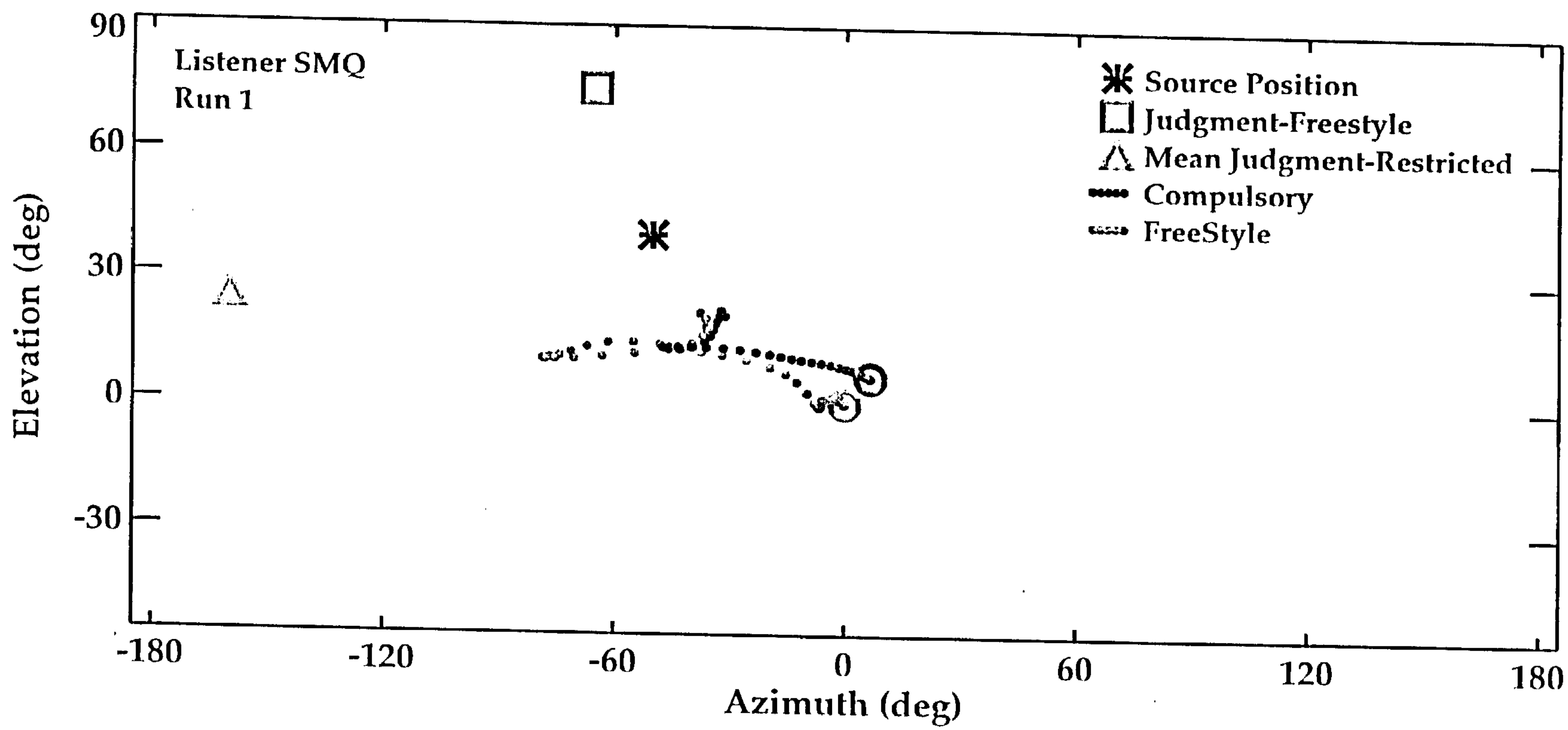
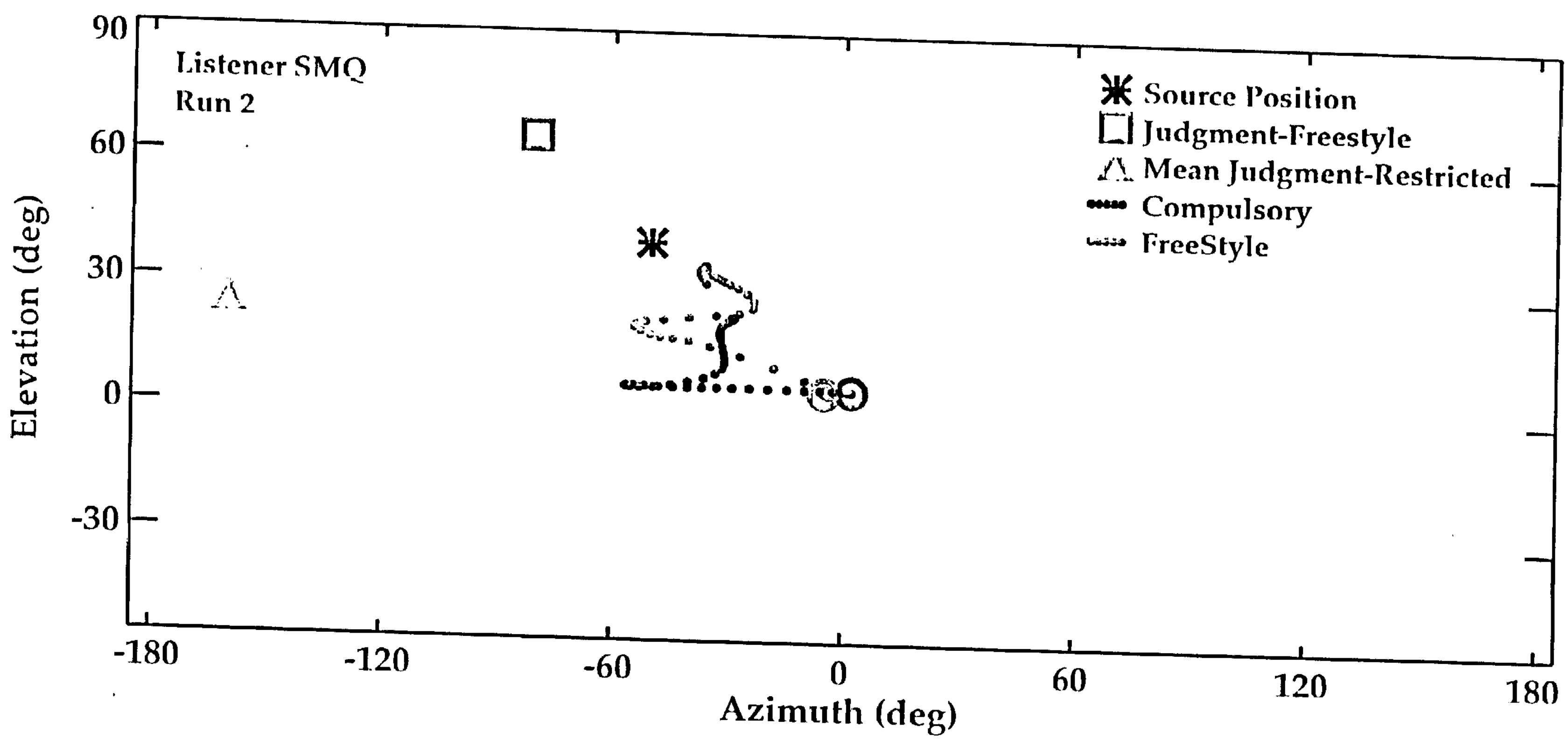


Figure 3a

3A.

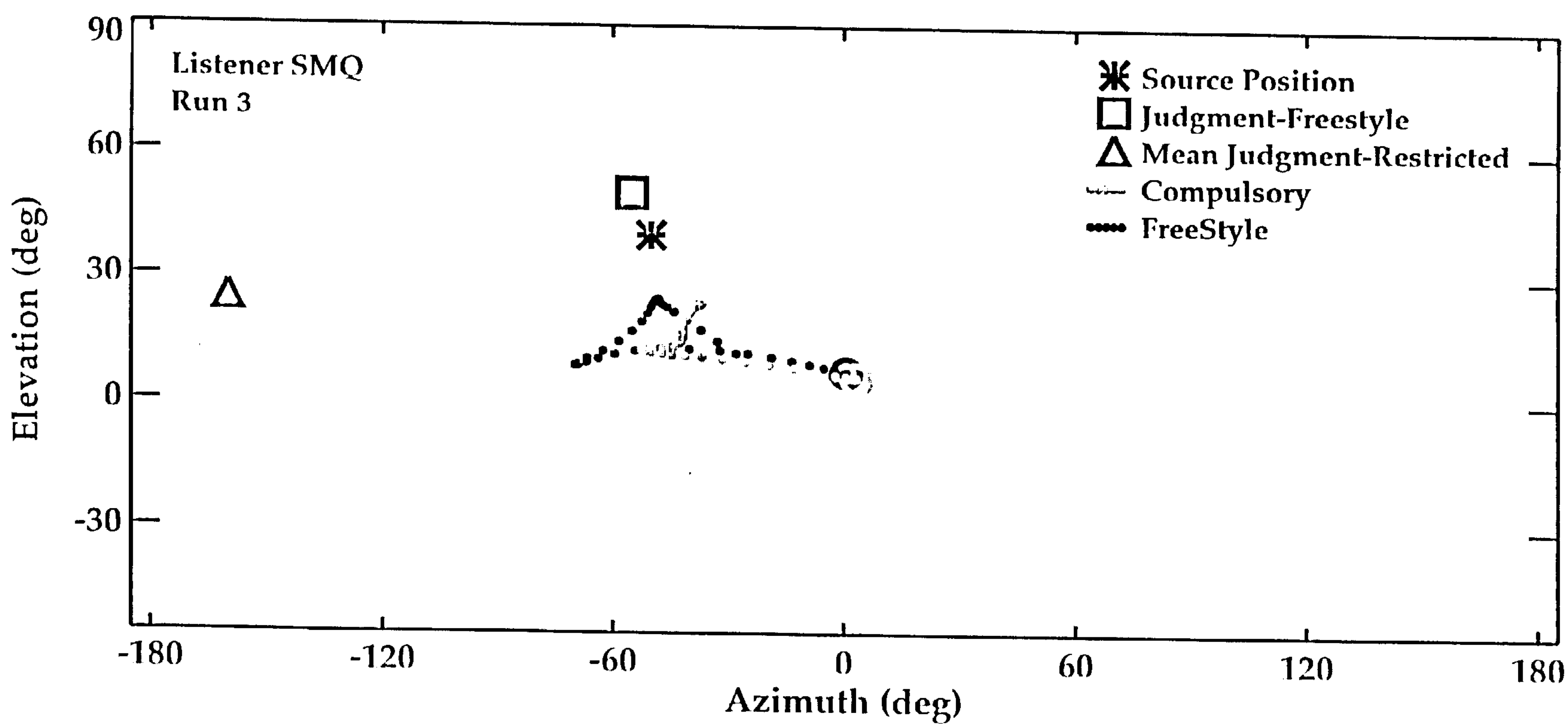
PRECEDING PAGE BLANK NOT FILMED



Listener SMQ
Run 2
Source Position
Judgment-Freestyle
Mean Judgment-Restricted
Compulsory
FreeStyle
Elevation (deg)
90
60
30
0
-30
-180
-120
-60
0
60
120
180
Azimuth (deg)

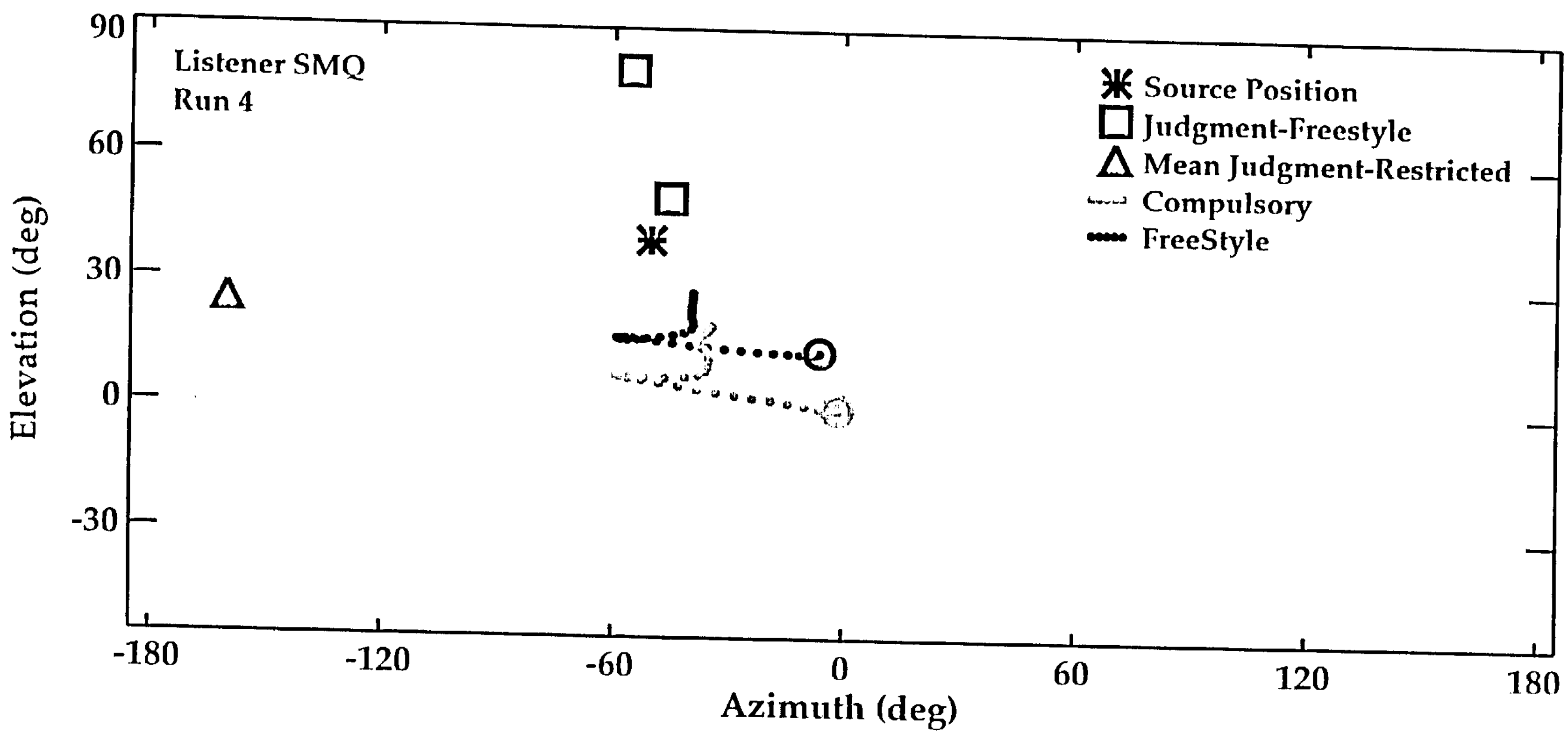
3B.

Figure 3b


Listener SMQ
Run 3
Source Position
Judgment-Freestyle
Mean Judgment-Restricted
Compulsory
FreeStyle
Elevation (deg)
90
60
30
0
-30
-180
-120
-60
0
60
120
180
Azimuth (deg)

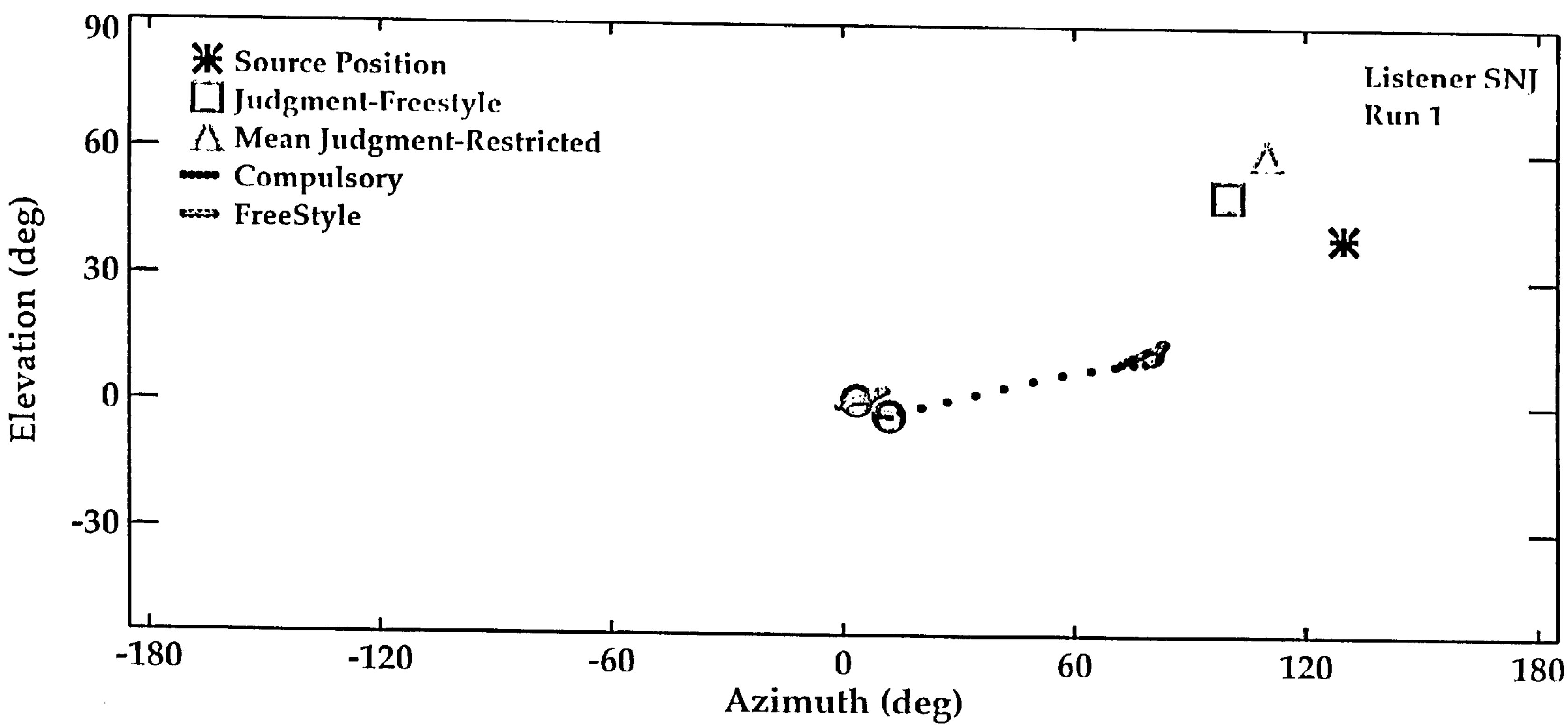
3c

Figure 3c


Listener SMQ
Run 4
Source Position
Judgment-Freestyle
Mean Judgment-Restricted
Compulsory
FreeStyle
Elevation (deg)
90
60
30
0
-30
Azimuth (deg)
-180
-120
-60
0
60
120
180

3D

Figure 3d

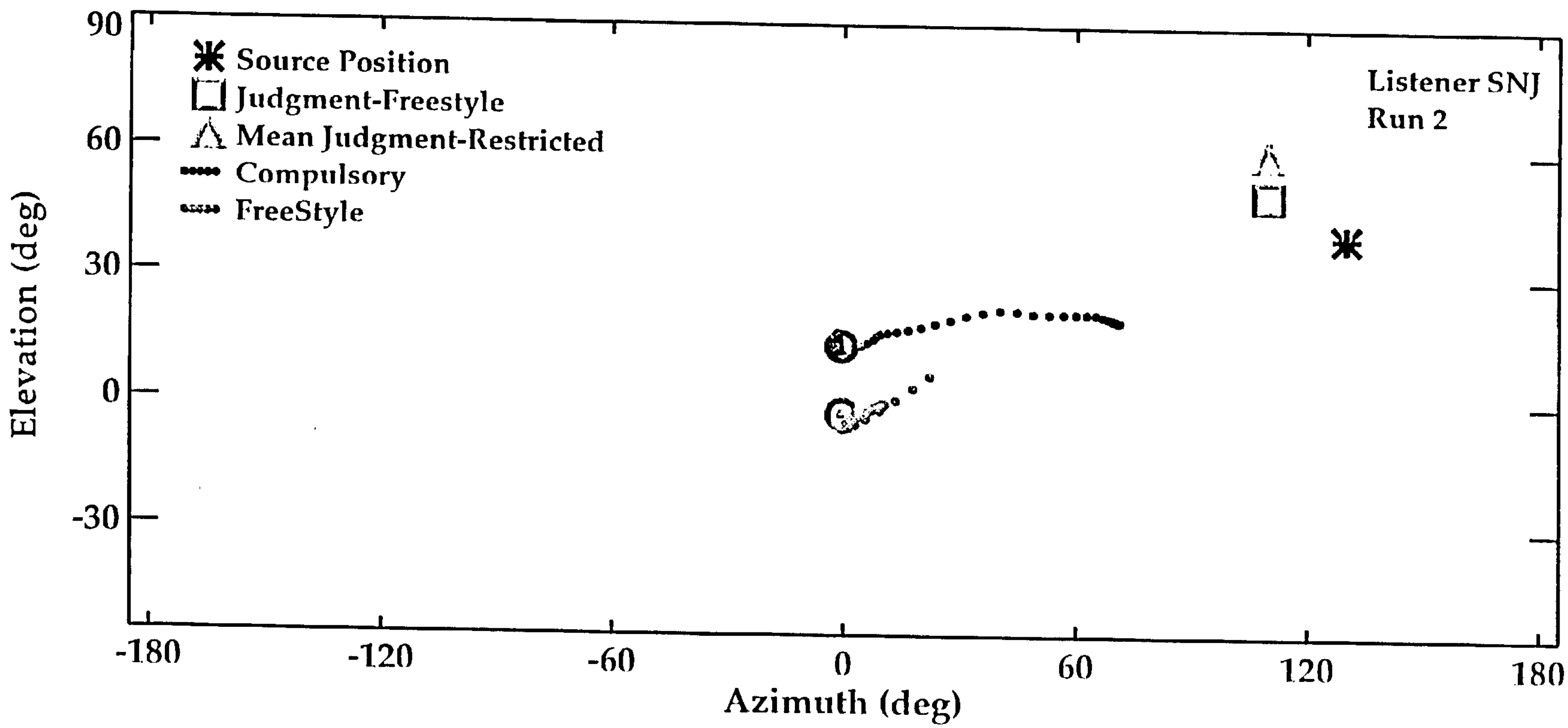


Source Position
Judgment-Freestyle
Mean Judgment-Restricted
Compulsory
FreeStyle
Listener SNJ
Run 1
Elevation (deg)
90
60
30
0
-30
-180
-120
-60
0
60
120
180
Azimuth (deg)

4A

Figure 4a

Figure 4b


Source Position
Judgment-Freestyle
Mean Judgment-Restricted
Compulsory
FreeStyle
Listener SNJ
Run 2
90
60
30
0
-30
Elevation (deg)
-180
-120
-60
0
60
120
180
Azimuth (deg)

4B